PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-288294
(43)Date of publication of application : 19.10.1999

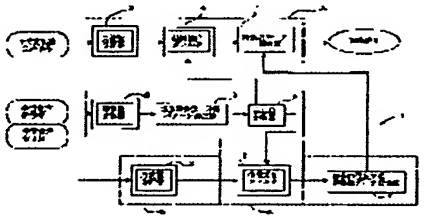
(51)Int.Cl. G10L 3/00

(21)Application number : 10-091116
(22)Date of filing : 03.04.1998

(71)Applicant : HONDA MOTOR CO LTD
(72)Inventor : AKATSUKA KOJI

(54) VOICE RECOGNITION DEVICE

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a voice recognition device which keeps a high performance even for a variety of speaking of unspecified speakers and to reduce erroneous recognition by a simple constitution.
SOLUTION: The voice recognition device is provided with a frequency analyzer 2 which successively obtains frequency spectrums, which are obtained by frequency analysis of a voice signal, along the time base to convert it into a time series data group, a partial frequency-time pattern generator 3 which segments output time series data from the frequency analysis means, to which the voice signal based on voices spoken by plural learning speakers is inputted, by a preliminarily determined time window, a main component analyzer 4 which uses a time series data group segmented by this pattern generator 3 to analyze a main component, and a feature extraction filter 5 which uses a lower-order main component obtained by main component analysis as the base to compress input time series data to lower-order time series data.



LEGAL STATUS

[Date of request for examination] 29.11.2004
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]
[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's decision of rejection]
[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号

特開平11－288294

(43)公開日　平成11年(1999)10月19日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ１０Ｌ　3/00	５１５	Ｇ１０Ｌ　3/00	５１５Ａ

審査請求　未請求　請求項の数２　ＯＬ　（全 11 頁）

(21)出願番号　特願平10－91116

(22)出願日　平成10年(1998)４月３日

(71)出願人　000005326
本田技研工業株式会社
東京都港区南青山二丁目１番１号

(72)発明者　赤塚　浩二
埼玉県和光市中央１丁目４番１号　株式会社本田技術研究所内

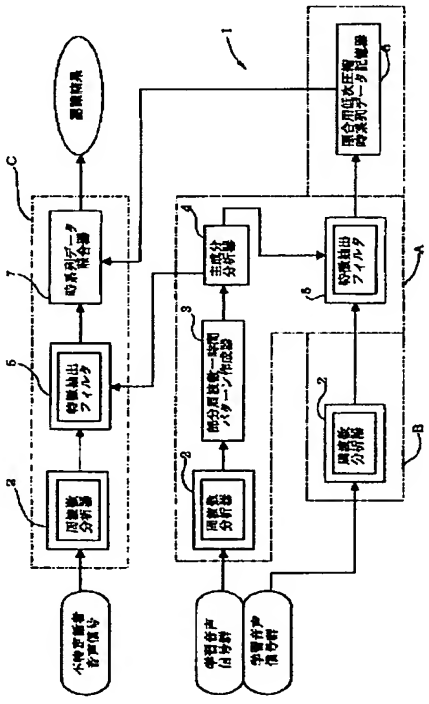
(74)代理人　弁理士　大西　正悟

(54)【発明の名称】　音声認識装置

(57)【要約】

【課題】　簡単な構成で、不特定話者の発話の多様性に対しても高性能を維持することができて、誤認識を低減させた音声認識装置を得る。

【解決手段】　音声信号を周波数分析して得た周波数スペクトルを、時間軸に沿って順次求めて時系列データ群に変換する周波数分析器２と、複数の学習話者から発話された音声に基づく音声信号が入力された周波数分析手段からの出力時系列データを予め定めた時間窓で切り出す部分周波数－時間パターン作成器３と、このパターン作成器３によって切り出された時系列データ群を用いて主成分分析を行う主成分分析器４と、主成分分析により得た低次の主成分を基底として入力時系列データを低次の時系列データに圧縮する特徴抽出フィルタ５とを備えて音声認識装置が構成される。



1

【特許請求の範囲】

【請求項１】　音声信号を周波数分析して得た周波数スペクトルを、時間軸に沿って順次求めて時系列データ群に変換する周波数分析手段と、

複数の学習話者から発話された音声に基づく音声信号が入力された前記周波数分析手段からの出力時系列データを予め定めた時間窓で切り出す切り出し手段と、

この切り出し手段によって切り出された時系列データ群を用いて主成分分析を行う主成分分析手段と、

前記主成分分析により得た低次の主成分を基底として入力時系列データを低次の時系列データに圧縮する特徴抽出フィルタ手段とを備え、

前記特徴抽出フィルタ手段に用いる前記基底は各主成分の時間窓の中央付近の周波数軸方向の成分で構成されるとともに、前記基底の時間軸方向の窓サイズはこれら各主成分の時間軸方向の幅よりも小さく、

前記複数の学習話者から発話された音声に基づく低次の時系列データと不特定話者から発話された音声に基づく低次の時系列データとを照合し、この照合結果に基づいて音声認識を行うことを特徴とする音声認識装置。

【請求項２】　前記周波数分析手段によって求められる出力時系列データの周波数軸が、メルスケール等の対数スケールで表示されることを特徴とする請求項１に記載の音声認識装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、不特定話者から離散的に発話された音声を自動的に認識する音声認識装置に関する。

【０００２】

【従来の技術】複数の不特定話者からの音声を誤認識せずに認識する従来の音声認識装置の多くは、種々の周波数分析手法を用いて音声信号に対してある程度の周波数解像度を有する周波数分析を行って周波数－時間の符号系列に変換し、出現が予想される音素の隠れマルコフモデルを用意し、さらにこのように容易した隠れマルコフモデルを多くの話者からの発話音声によって学習させて予め用意しておく。

【０００３】この学習済みの隠れマルコフモデルを用いて、不特定話者から発話された音声に基づく周波数－時間の符号系列の部分区間を、すべての音素モデルと照合することによって音素系列の候補の時系列に変換し、この音素の時系列が最も良く表される単語を認識結果として出力するようになされている。

【０００４】

【発明が解決しようとする課題】しかしながら、従来の音声認識装置では、不特定話者の発話の多様性に対応して高性能な音声認識特性を維持するための隠れマルコフモデルの学習に多くの学習データを必要とし、隠れマルコフモデルで音素を精密に特定するためにある程度の周

2

波数分析の解像度、すなわち、ある程度の大きさのベクトル次数を必要とするという問題があった。

【０００５】この結果、隠れマルコフモデルの学習時と音素特定時の演算負荷が重く、さらに単語の認識過程に置いて少なくとも音素照合と単語照合の二段階の照合演算処理を必要とするという問題点があった。

【０００６】本発明は、簡単な構成で、不特定話者の発話の多様性に対しても高性能を維持することができて、誤認識を低減させた音声認識装置を提供することを目的とする。

【０００７】

【課題を解決するための手段】このようなことから本発明に係る音声認識装置は、音声信号を周波数分析して得た周波数スペクトルを、時間軸に沿って順次求めて時系列データ群に変換する周波数分析手段（例えば、図１における周波数分析器２）と、複数の学習話者から発話された音声に基づく音声信号が入力された前記周波数分析手段からの出力時系列データを予め定めた時間窓で切り出す切り出し手段（例えば、図１における部分周波数－時間パターン作成器３）と、この切り出し手段によって切り出された時系列データ群を用いて主成分分析を行う主成分分析手段（例えば、図１における主成分分析器４）と、主成分分析により得た低次の主成分を基底として入力時系列データを低次の時系列データに圧縮する特徴抽出フィルタ手段（例えば、図１における特徴抽出フィルタ５）とを備えて構成される。

【０００８】さらに、この音声認識装置では、特徴抽出フィルタ手段に用いる基底は、各主成分の時間窓の中央付近の周波数軸方向の成分で構成されるとともに、この基底の時間軸方向の窓サイズはこれら各主成分の時間軸方向の幅よりも小さく、複数の学習話者から発話された音声に基づく低次の時系列データと不特定話者から発話された音声に基づく低次の時系列データとを照合し、この照合結果に基づいて音声認識を行う。ここで、主成分のうち低次主成分は音声の識別的特徴に多く含まれる成分の固有空間を定義しており、時系列データに基づく周波数－時間パターンの中に最も多く含まれる部分の特徴を表しているので、音声信号に含まれる学習話者の個人性に基づく成分や認識に悪影響を及ぼすと考えられるノイズ成分は低次成分に含まれず、音声認識が正確となる。

【０００９】また、特徴抽出フィルタ手段に用いる基底の時間軸方向の窓サイズは各主成分の時間軸方向の幅よりも小さく設定されるため、音声信号の音素を区切るラベル位置の精度が多少悪くてもこれを吸収できる。さらに、各音素の特徴は時間軸方向における各音素の中央部に存在する可能性が高いため、時間窓の中央が各音素の中央に一致するように設定すれば、情報の欠落を最小限に抑えることができ、より正確な音声認識が可能となる。

【００１０】なお、周波数分析手段によって求められる出力時系列データの周波数軸を、メルスケール等の対数スケールで表示するのが好ましい。一般的に、音声における高い周波数領域では話者の違いによる特徴の変動が大きく、低い周波数領域ではこの変動が小さい。周波数軸を対数スケールとすれば、高い周波数領域における話者の違いよる特徴変動を抑えるとともに低い周波数領域での特徴を大きくすることにより、話者の違いによる特徴変動に対して、音声認識装置が鈍感となり、話者の相違に対して認識率を向上させることができる。

【００１１】

【発明の実施の形態】以下、本発明に係る音声認識装置の好ましい実施形態について図面を参照して以下に説明する。図１に本発明の一実施形態に係る音声認識装置の構成を模式ブロック図を用いて示している。この図において、作用の理解を容易にするために、同一の構成要素であっても異なる音声信号ラインに使用する構成要素は重複して示している。図１において二重枠の構成要素がこれに当たり、同一符号は同一の構成手段であることを示している。

【００１２】この音声認識装置１は、複数の学習話者から発せられる発話音声に基づき学習話者の音素に対する特徴を抽出し、抽出した特徴を基底とする特徴抽出フィルタを作成する特徴抽出フィルタ作成部Ａと、複数の学習話者の発話、例えば単語の音声信号に基づく情報を特徴抽出フィルタに供給し、特徴抽出フィルタによって前記情報を圧縮して照合用低次圧縮時系列データ群を生成する照合時系列データ作成部Ｂと、入力された不特定話者からの音声信号を特徴抽出フィルタに供給して、特徴抽出フィルタによって圧縮した時系列データを生成し、この時系列データを照合用低次圧縮時系列データと照合して音声認識結果を出力する不特定話者音声認識部Ｃとを備えている。

【００１３】特徴抽出フィルタ作成部Ａは、複数の学習話者から発話された音声（以下、学習音声群とも称す）の周波数スペクトルの時間的変化を示すため、複数の学習話者から発話された音声に基づく音声信号を周波数分析して得た周波数スペクトルを、時間軸に沿って順次求めた時系列データ群（周波数－時間の時系列データ群）に変換する周波数分析器２と、周波数分析器２によって変換された前記複数の学習話者からの音声に基づく周波数－時間の時系列データ群から小さな時間窓の範囲における部分周波数－時間の時系列データを切り出す部分周波数－時間パターン作成器３と、部分周波数－時間パターン生成器３によって切り出された複数の部分周波数－時間の時系列データを用いて主成分分析を行う主成分分析器４と、主成分分析器４による主成分分析結果の低次主成分を基底にする特徴抽出フィルタ５とを備えて、複数の学習話者から学習話者の音素に対する特徴を抽出する。

【００１４】照合時系列データ作成部Ｂは照合用低次圧縮時系列データ記憶器６を備え、複数の学習話者から発話された単語音声の周波数スペクトルの時間的変化を示すため、複数の学習話者から発話された前記単語音声の音声信号を周波数分析器２によって周波数分析して得た周波数スペクトルを、時間軸に沿って順次求めた周波数－時間の時系列データ群に変換し、変換された周波数－時間の時系列データ群を特徴抽出フィルタ５に送出し、特徴抽出フィルタ５にて周波数－時間の時系列データを次元圧縮して照合用低次圧縮時系列データ群を得て、照合用低次圧縮時系列データ記憶器６に記憶させる。

【００１５】不特定話者音声認識部Ｃは時系列データ照合器７を備え、不特定話者から発話された音声の周波数スペクトルの時間的変化を示すため、不特定話者から発話された音声に基づく音声信号を周波数分析器２によって周波数分析して得た周波数スペクトルを、時間軸に沿って順次求めた周波数－時間の時系列データ群に変換し、変換された周波数－時間の時系列データ群を特徴抽出フィルタ５に送出し、特徴抽出フィルタ５にて周波数－時間の時系列データを次元圧縮して時系列データ群を得て、時系列データ群と照合用低次圧縮時系列データ記憶器６から読み出した照合用低次圧縮時系列データとを時系列データ照合器７にて照合し、照合用低次圧縮時系列データ群中から、時系列データ群に最も近いものを求め、照合結果に基づいて不特定話者からの発声音声に基づく言葉を認識する。

【００１６】次に、周波数分析器２、部分周波数－時間パターン作成器３、主成分分析器４、特徴抽出フィルタ５のそれぞれについて具体的に説明する。

【００１７】周波数分析器２では、入力信号がＡ／Ｄ変換され、Ａ／Ｄ変換された音声信号に対して、高域強調処理がなされ、高域強調処理されたＡ／Ｄ変換音声信号に対して時間窓としてハニング窓をかけて、短時間の音声信号を切り出し、切り出した短時間音声信号をフーリエ変換を行うことで、周波数展開を行い、線形の周波数軸を対数尺度に近いメルスケールに変換する。この処理を時間軸に沿って繰り返すことで、音声スペクトルの時間的変化を示すための周波数－時間の時系列データに変換される。したがって、周波数分析器２では、入力音声のサウンドスペクトルパターンに実質的に展開される。以下、この周波数－時間の時系列データの周波数軸方向の点数をＮで表すことにする。

【００１８】この周波数分析手法に応じて特徴抽出フィルタ５を作成すれば、音声情報の欠落が少ない。また、周波数分析に応じて特徴抽出フィルタ５を作成したときに音声情報に欠落がないような他の周波数分析手法によっても良い。従って、周波数分析器２による方法によれば、さらにベクトル次数の少ない周波数－時間パターンやケプストラム等にも適用することができる。この結果、周波数－時間の時系列データ群によって実質的に音

声信号の周波数－時間パターンが示される。

【００１９】部分周波数－時間パターン作成器３では、周波数分析器２から出力される周波数－時間の時系列データ群中から、所定の小さな時間窓の範囲における周波数－時間の時系列データが切り出される。このため、部分周波数－時間パターン作成器３から出力される周波数－時間の時系列データに基づく音声の周波数－時間パターンは、周波数分析器２から出力される周波数－時間の時系列データに基づく音声の周波数－時間パターンの一部分であって、部分周波数－時間パターンであるといえる。

【００２０】特徴抽出フィルタ５は、周波数－時間の時系列データからの情報の欠落を最小限に抑え、情報圧縮した時系列データを作成する。本例では情報の圧縮に主成分分析を用いている。

【００２１】さらに詳細に、例えば、９名の異なる学習話者の共通した１００語の発話データを学習音声信号群として用いた場合の例を説明する。

【００２２】この場合、会話データには、単語音声信号区間中の発話音素と、発話音素の音声信号の時間軸上における開始点と終了点とに対応が付けられたラベルデータとが予め設定されている。例えば、図３（Ａ）に示すように、音素Ｅに対する開始点の時間ラベルａ、音素Ｅに対する終了点の時間ラベルであり且つ音素Ｆに対する開始点の時間ラベルである時間ラベルｂ、音素Ｆに対する終了点の時間ラベルｃを持っている。なお、図３（Ａ）における横軸は時間で、縦軸が周波数であり、各周波数の強度スペクトルが紙面に垂直な値で表され、いわゆる三次元グラフとなるデータを構成している。

【００２３】部分周波数－時間パターン作成器３は、周波数分析器２から出力される周波数－時間の時系列データをラベルデータとともに、時間軸上の音素の中心位置、図３（Ａ）に示す例では（ａ＋ｂ）／２、（ｂ＋ｃ）／２を求め、この中心位置を中心に時間窓部分の周波数－時間の時系列データを切り出す。

【００２４】すなわち、学習音声信号群に対して、部分周波数－時間パターン作成器３によって、例えば、３０ｍｓの時間窓Ｄで切り出しを行い、部分周波数－時間の時系列データ群を作成する。部分周波数－時間パターン作成器３によって作成された部分周波数－時間の時系列データの時間窓Ｄによる切り出しは、図３（Ｂ）に示すように、音素Ｅに対しては時間ラベルａと時間ラベルｂとの間の中央に時間窓Ｄが来るように、［｛（ａ＋ｂ）／２｝－（Ｄ／２）］の位置から［｛（ａ＋ｂ）／２＋（Ｄ／２）］の位置までが切り出され、音素Ｅに対しては時間ラベルｂと時間ラベルｃとの中央に時間窓Ｄが来るように、［｛（ｂ＋ｃ）／２｝－（Ｄ／２）］の位置から［｛（ｂ＋ｃ）／２＋（Ｄ／２）］の位置までが切り出される。

【００２５】この切り出し処理を同じ音素のラベル区間について行うことによって、同じ音素の周波数－時間の時系列データを複数集めることができる。同じ音素を複数集めた周波数－時間の時系列データの平均値を求め、これを部分周波数－時間の時系列データとする。この部分周波数－時間の時系列データを音素毎に作成することによって部分周波数－時間の時系列データ群が作成される。この部分周波数－時間の時系列データ群の作成処理により、このように各音素の時間長さより短い時間窓による切り出しを行えば、各音素のラベル区間のラベル時刻の精度の悪さを吸収できる。また、音素のラベル区間における音素毎の特徴は、ラベル区間のほぼ中央に存在する可能性が高いため、開始および終了ラベルの中央に時間窓の中心が位置するようにして切り出しを行うことにより情報の欠落を最小限に抑えることができる。

【００２６】この時間窓による切り出し処理を、時間軸方向の特徴変化の少ない音素毎、すなわち、比較的定常的な音素毎に行っても良い。

【００２７】この部分周波数－時間の時系列データ群から、主成分分析器４によって主成分が求められるが、これについて図４に基づいて説明する。図４においては、部分周波数－時間の時系列データをパターンと略記してある。

【００２８】切り出された音素Ａの部分周波数－時間の時系列データ群、音素Ｂの部分周波数－時間の時系列データ群、・・・、音素Ｚの部分周波数－時間の時系列データ群は図４（Ａ）に模式的に示すように発話データに含まれる各音素のパターンからなり、それぞれ複数のパターンを有している。そして、各音素Ａ～Ｚについての部分周波数－時間の時系列データ群の平均値が求められる。その結果、音素Ａの部分周波数－時間の時系列データ群の平均値、音素Ｂの部分周波数－時間の時系列データ群の平均値、・・・、音素Ｚの部分周波数－時間の時系列データ群の平均値が、図４（Ｂ）に模式的に示す如く得られる。

【００２９】各音素Ａ～Ｚの部分周波数－時間の時系列データの平均値は主成分分析器４によって、図４（Ｃ）に模式的に示すように、主成分分析が行われる。主成分分析の結果、図４（Ｄ）に模式的に示すように、第１主成分、第２主成分、・・・、第Ｋ主成分が求められる。主成分を求める場合のサンプルデータ数は、そのサンプルデータを定義するベクトル次数より多く必要である。したがって、音素Ａから音素Ｚの個数が、部分周波数－時間の時系列データの次元数よりも少ない場合、各音素毎に求めた平均値に近い部分周波数－時間の時系列データを数個ずつ求め、これを図４（Ｂ）に示す各音素のパターンの平均値の代わりに用いても良い。

【００３０】すなわち、主成分分析ではサンプルデータ空間のベクトル次元数と同数の次元数の主成分が求められ、サンプルデータの分散が最も多い軸を決める主成分を第１主成分、分散が２番目に大きい軸を決める主成分

を第２主成分、以下同様に第Ｋ主成分が決まる。

【００３１】主成分分析器４では分散の大きい第１主成分から順次分散が減少する第５番目の主成分を低次主成分として用いている。すなわち、情報の損失量の最小から最大の方向へ五つの主成分を低次主成分として用いる。従って、主成分のうちの低次主成分は部分周波数－時間の時系列データ群の特徴に多く含まれる成分の固有空間を定義しており、音声信号の周波数－時間の時系列データに基づく周波数－時間パターンの中に最も含まれる部分の特徴を表している。すなわち、音声信号に含まれる学習話者の個人性に基づく成分や認識に悪影響を及ぼすと考えられるノイズ成分は、低次主成分には含まれていないと考えられる。

【００３２】特徴抽出フィルタ５では、この低次主成分を基底として用いて、例えば五つの第１～第５低次主成分ベクトルδ１ｉ～δ５ｉを特徴抽出フィルタ５の基底として用い、周波数分析器２から出力される周波数－時間の時系列データの各時刻における周波数－時間の時系列データと、第１～第５低次主成分ベクトルδ１ｉ～δ５ｉとの間で相関値を求める。この各低次主成分毎の相関値出力をチャンネルとも称する。この相関値を各チャンネル毎に正規化して、五つのチャンネルのフィルタ出力とする。

【００３３】上記からも明らかなように、特徴抽出フィルタ５は五つの低次主成分の場合を例に示せば、図２に示すように、時間窓幅点数ｄｔの周波数分析結果のＮ×ｄｔ次元ベクトルＸｉと各低次主成分ベクトルδ１ｉ～δ５ｉとの積和演算を各時刻において積和演算器５１１～５１５にてそれぞれ入力Ｎ×ｄｔ次元ベクトルに対して行って、各積和演算器５１１～５１５からの出力を、正規化器５２１～５２５によってそれぞれレベルを正規化して、正規化された各正規化器５２１～５２５からの出力を各チャンネルの出力として送出する。

【００３４】次に、照合用低次圧縮時系列データ群の作成について説明する。各単語の学習音声信号が周波数分析器２に供給されて、学習音声信号に基づく周波数－時間の時系列データが作成される。この周波数－時間の時系列データが既に学習音声信号群における音素に対して求めておいた低次主成分を基底とする特徴抽出フィルタ５に供給され、特徴抽出フィルタ５において次元圧縮されて特徴抽出フィルタ５の各チャンネルから時系列データが出力され、この時系列データが照合用低次圧縮時系列データとされる。

【００３５】このように作成された照合用低次圧縮時系列データの構造は、図５に示すように構成され、それぞれ学習音声の発話者による同じ単語の学習音声による場合の照合用低次圧縮時系列データであり、９名の話者による１００単語に対する場合には９００個の照合用低次圧縮時系列データ群が得られ、照合用低次圧縮時系列データ群の各要素は学習音声信号の各発話単語名とそれに対応する照合用低次圧縮時系列データの対で構成される。この照合用低次圧縮時系列データ群は照合用低次圧縮時系列データ記憶器６に記憶される。

【００３６】上記のように照合用低次圧縮時系列データが照合用低次圧縮時系列データ記憶器６に記憶させてある状態で、不特定話者からの音声認識が行われる。不特定話者からの入力音声信号は周波数分析器２によって周波数分析され、既に学習音声信号群からの音声信号に基づいて予め特徴抽出フィルタ作成部Ａで求められた特徴抽出フィルタ５に供給されて、特徴抽出フィルタ５において次元圧縮処理がなされて、時系列データに変換される。

【００３７】不特定話者からの音声信号に基づく時系列データは、学習音声信号群に基づいて照合時系列データ作成部Ｂで求められた照合用低次圧縮時系列データ群との間で時系列データ照合器７において照合されて、不特定話者からの音声信号に基づく時系列データに最も近い照合用低次圧縮時系列データが照合用低次圧縮時系列データ群中から選び出され、選び出された照合用低次圧縮時系列データに対する発話単語名が認識結果として出力される。

【００３８】次に、本実施形態における時系列データ照合器７をＤＰ（dynamic programming)法を用いた照合の場合を例に説明する。

【００３９】ＤＰ法は、入力時系列データと予め記憶された時系列データ群の間で、非線形に時間伸縮することで時間正規化を行い対応付けを行う照合法である。この方法によれば、入力時系列データと予め記憶された各時系列データの間の時間正規化後の距離が定義され、この距離が最小である時系列データが入力時系列データを最も良く表すものとし、認識結果とするものである。本実施形態では、このＤＰ法が不特定話者からの音声信号に基づく時系列データと照合用低次圧縮時系列データとの間に適用されて、時間正規化後の最小距離を持つ照合用低次圧縮時系列データに対応させた単語名が出力される。

【００４０】次に本実施の形態に基づく評価実験結果について説明する。ここではテストサンプルとして、話者１０名分の４９２単語の離散発生単語データベースを用いて、この内の１００単語及び４９２単語を用いた場合の評価結果について、以下に記す。

【００４１】最初、評価単語数を１００単語にした場合の評価結果について記載する。テスト話者１名を除く９名の話者の発話データを学習音声信号群として用いて特徴抽出フィルタ作成部Ａで特徴抽出フィルタ５を作成した。サンプルとして用いた音素は母音、破裂音、摩擦音、鼻音であり、部分周波数－時間パターン作成器３を用いて、話者毎に部分周波数－時間の時系列データを求め、この部分周波数－時間の時系列データから主成分分析器３で主成分を求め、この主成分のうち、低次主成分

の第８主成分までを用いた。

【００４２】時系列データ照合器７で用いる照合用低次圧縮時系列データ群は、前記テスト話者１名を除く９名の話者の発話データを学習音声信号群として、上記特徴抽出フィルタ５を用いた照合時系列データ作成部Ｂで９００個の照合用低次圧縮時系列データを求めた。評価実験では、テスト話者を変えながら行い、その都度、特徴抽出フィルタ５を求め直し、照合用低次圧縮時系列データを作成し直した。

【００４３】一方、特徴抽出フィルター５の出力チャンネル数は２から８間で変化させた。周波数軸点数Ｎを６４，３０ｍｓに相当する時間窓幅点数ｄｔを６に設定した場合の認識結果を図６に示す。周波数軸点数Ｎを６４，特徴抽出フィルターに用いる規定を時間窓の中央付近の時間窓幅点数ｄｔを１に設定した場合の認識結果を図７に示す。いずれの手法でも、特徴抽出フィルターのチャンネル数を５チャンネルに設定した場合、どの話者に対しても、９６％以上の認識率であった。特徴抽出フィルターに用いる規定の時間窓幅点数ｄｔを１に設定した場合、積和演算の計算負荷は１／６倍に軽減されるが、それでも、認識性能は同等維持できる。周波数軸点数Ｎを３２、時間窓幅点数ｄｔを６に設定した場合の認識結果を図８に示す。周波数軸点数Ｎを３２、特徴抽出フィルターに用いる基底を時間窓の中央付近の時間窓幅点数ｄｔを１に設定した場合の認識結果を図９に示す。周波数軸点数Ｎを３２に設定した場合でも、５チャンネルに設定した場合、どの話者に対しても、認識率９５％以上を確保している。

【００４４】次に、評価単語数を４９２単語にした場合の評価結果について、周波数軸点数Ｎを６４、５ｍｓに相当する時間窓幅点数ｄｔを１、特徴抽出フィルターのチャンネル数を５チャンネルに設定した場合の認識結果を図１０に示す。どの話者に対しても９０％以上の認識率、話者平均の認識率が９４．６７％と、本手法は、語彙数増に対してもある程度の認識性能を維持できた。

【００４５】

【発明の効果】以上説明したように、本発明によれば、特徴抽出のための演算も、且つ照合のための処理も簡単なため、その構成は簡単であり、不特定話者の発話に対しても誤認識が少なく、音声認識をすることができるという効果が得られる。さらに、本発明の装置では、特徴抽出フィルタ手段に用いる基底は、各主成分の時間窓の中央付近の周波数軸方向の成分で構成されるとともに、この基底の時間軸方向の窓サイズはこれら各主成分の時間軸方向の幅よりも小さく、複数の学習話者から発話された音声に基づく低次の時系列データと不特定話者から発話された音声に基づく低次の時系列データとを照合し、この照合結果に基づいて音声認識を行う。ここで、主成分のうち低次主成分は時系列データ群の特徴に多く含まれる成分の固有空間を定義しており、時系列データに基づく周波数－時間パターンの中に最も多く含まれる部分の特徴を表しているので、音声信号に含まれる学習話者の個人性に基づく成分や認識に悪影響を及ぼすと考えられるノイズ成分は低次成分に含まれず、音声認識が正確となる。

【００４６】また、特徴抽出フィルタ手段に用いる基底の時間軸方向の窓サイズは各主成分の時間軸方向の幅よりも小さく設定されるため、音声信号の音素を区切るラベル位置の精度が多少悪くてもこれを吸収できる。さらに、各音素の特徴は時間軸方向における各音素の中央部に存在する可能性が高いため、時間窓の中央が各音素の中央に一致するように設定すれば、情報の欠落を最小限に抑えることができ、より正確な音声認識が可能となる。

【００４７】なお、周波数分析手段によって求められる出力時系列データの周波数軸を、メルスケール等の対数スケールで表示するのが好ましい。一般的に、音声における高い周波数領域では話者の違いによる特徴の変動が大きく、低い周波数領域ではこの変動が小さい。周波数軸を対数スケールとすれば、高い周波数領域における話者の違いよる特徴変動を抑えるとともに低い周波数領域での特徴を大きくすることにより、話者の違いによる特徴変動に対して、音声認識装置が鈍感となり、話者の相違に対して認識率を向上させることができる。

【図面の簡単な説明】

【図１】本発明の一実施形態に係る音声認識装置の構成を示す模式ブロック図である。

【図２】本発明の一実施形態に係る音声認識装置における特徴抽出フィルタの構成を示すブロック図である。

【図３】本発明の一実施形態に係る音声認識装置における部分周波数－時間パターン作成器の作用の説明に供する模式図である。

【図４】本発明の一実施形態に係る音声認識装置における部分周波数－時間パターン作成器および主成分分析器の作用の説明に供する模式図である。

【図５】本発明の一実施形態に係る音声認識装置における照合用低次圧縮時系列データの構造の一例を示す模式図である。

【図６】本発明の一実施形態に係る音声認識装置による音声認識結果（認識率）を示すグラフである。

【図７】本発明の一実施形態に係る音声認識装置による音声認識結果（認識率）を示すグラフである。

【図８】本発明の一実施形態に係る音声認識装置による音声認識結果（認識率）を示すグラフである。

【図９】本発明の一実施形態に係る音声認識装置による音声認識結果（認識率）を示すグラフである。

【図１０】本発明の一実施形態に係る音声認識装置による音声認識結果（認識率）を示すグラフである。

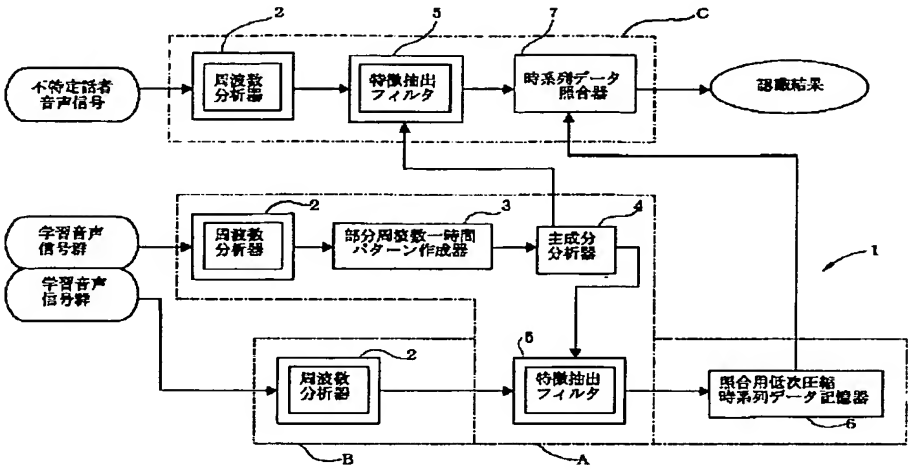
【符号の説明】

Ａ　特徴抽出フィルタ作成部

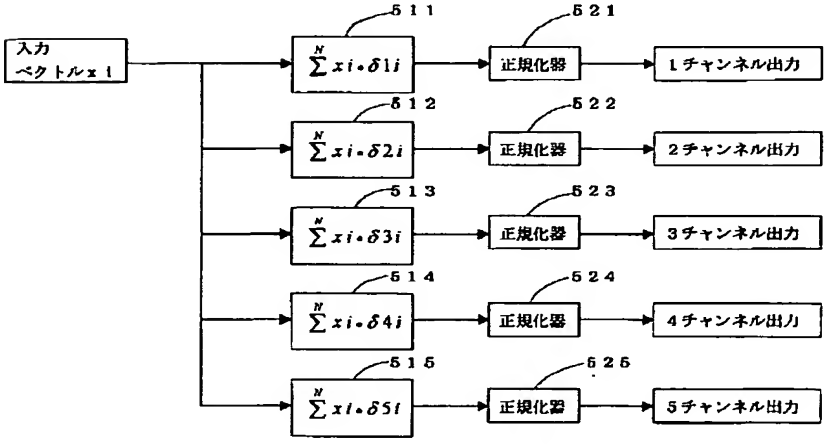
11　　12

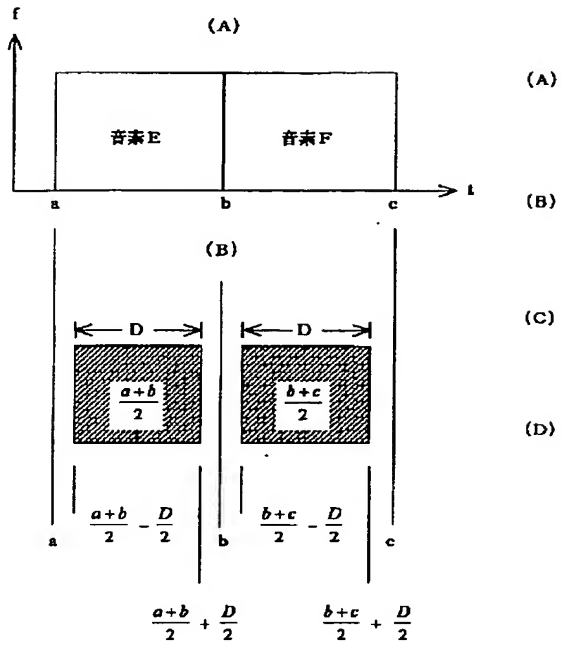
B　照合時系列データ作成部
C　不特定話者音声認識部
1　音声認識装置
2　周波数分析器（周波数分析手段）
3　部分周波数－時間パターン作成器（切り出し手段）＊
＊4　主成分分析器（主成分分析手段）
5　特徴抽出フィルタ（特徴抽出フィルタ手段）
6　照合用低次圧縮時系列データ記憶器
7　時系列データ照合器

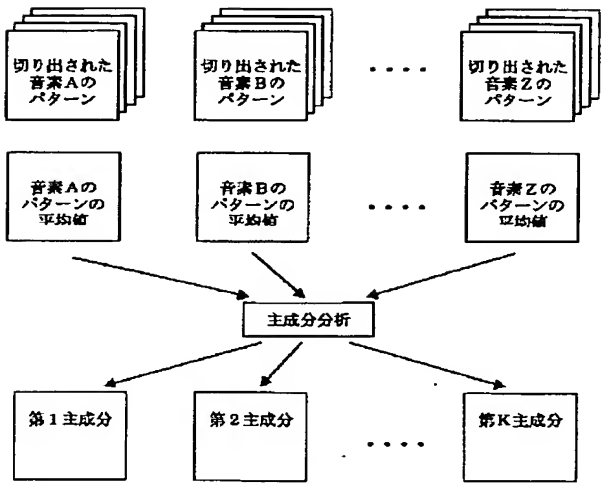
【図１】



【図２】







【図３】

【図４】

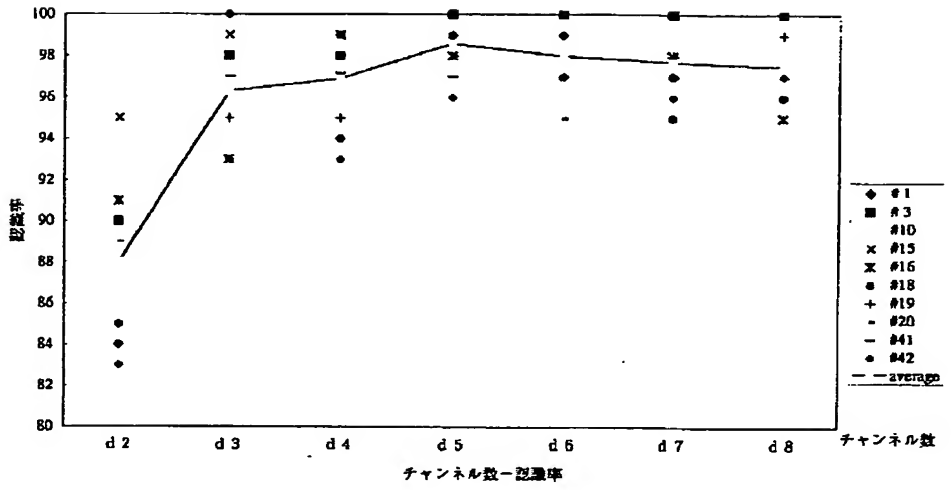
【図５】

単語コード	特徴フィルタの１チャンネルの時系列出力値
	特徴フィルタの２チャンネルの時系列出力値
	特徴フィルタの３チャンネルの時系列出力値
	特徴フィルタの４チャンネルの時系列出力値
	特徴フィルタの５チャンネルの時系列出力値

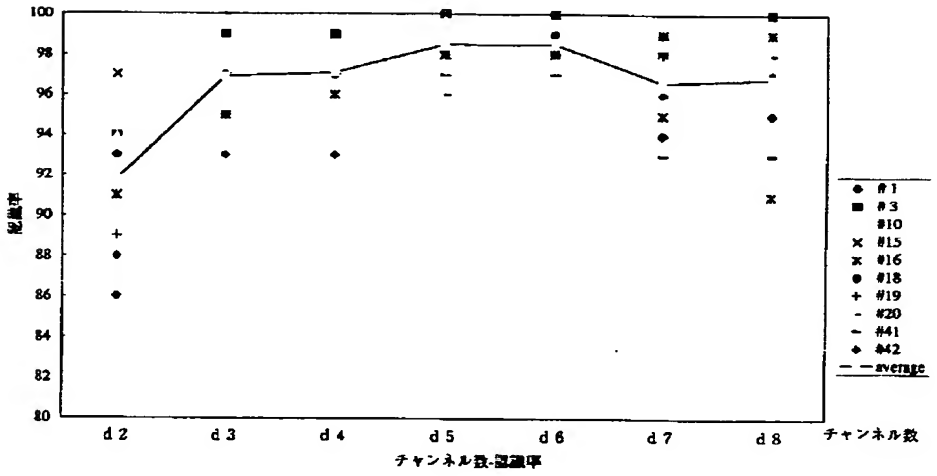
単語コード	特徴フィルタの１チャンネルの時系列出力値
	特徴フィルタの２チャンネルの時系列出力値
	特徴フィルタの３チャンネルの時系列出力値
	特徴フィルタの４チャンネルの時系列出力値
	特徴フィルタの５チャンネルの時系列出力値

単語コード	特徴フィルタの１チャンネルの時系列出力値
	特徴フィルタの２チャンネルの時系列出力値
	特徴フィルタの３チャンネルの時系列出力値
	特徴フィルタの４チャンネルの時系列出力値
	特徴フィルタの５チャンネルの時系列出力値

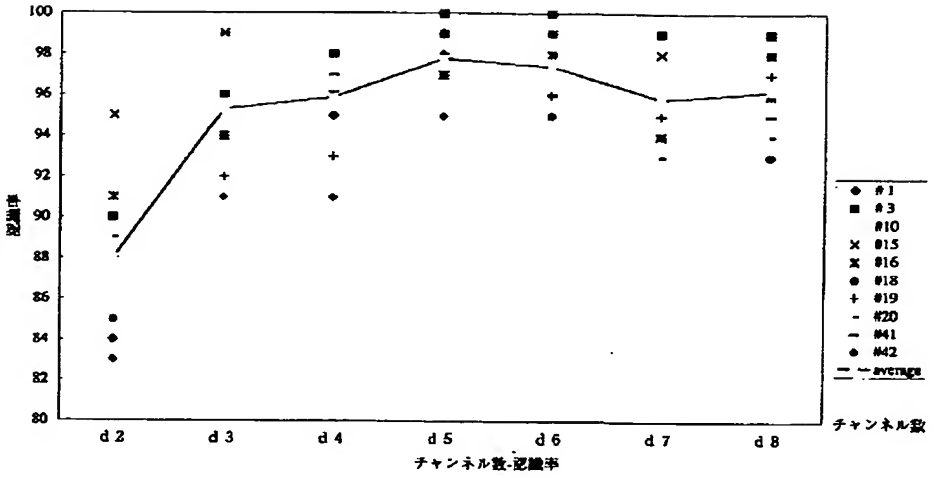
【図６】


100
98
96
94
92
90
88
86
84
82
80
認識率
d 2
d 3
d 4
d 5
d 6
d 7
d 8
チャンネル数
チャンネル数－認識率
#1
#3
#10
#15
#16
#18
#19
#20
#41
#42
average

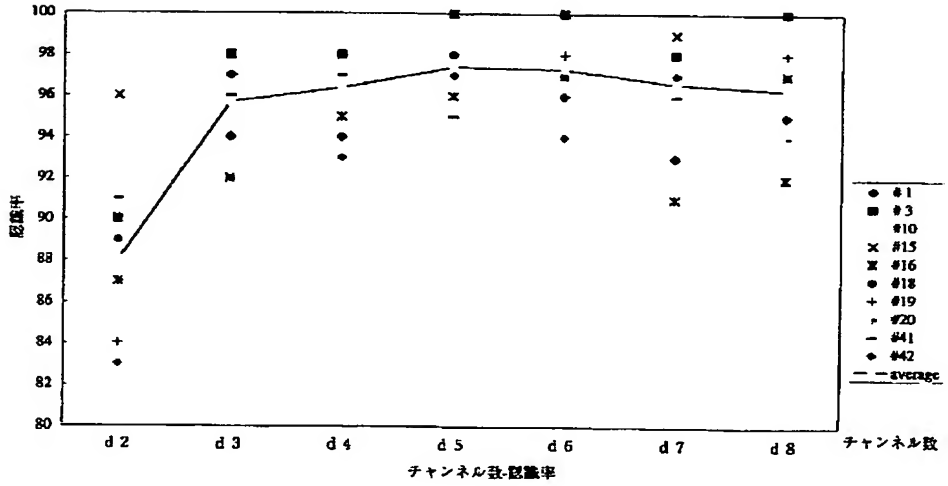
【図７】


100
98
96
94
92
90
88
86
84
82
80
認識率
d 2
d 3
d 4
d 5
d 6
d 7
d 8
チャンネル数
チャンネル数-認識率
#1
#3
#10
#15
#16
#18
#19
#20
#41
#42
average

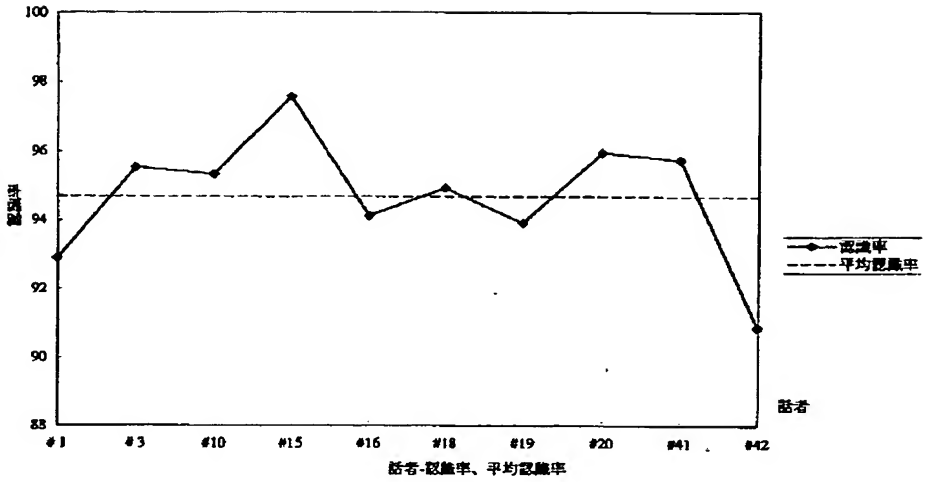
【図８】


100
98
96
94
92
90
88
86
84
82
80
認識率
d 2
d 3
d 4
d 5
d 6
d 7
d 8
チャンネル数-認識率
#1
#3
#10
#15
#16
#18
#19
#20
#41
#42
average
チャンネル数

【図９】


100
98
96
94
92
90
88
86
84
82
80
認識率
d 2
d 3
d 4
d 5
d 6
d 7
d 8
チャンネル数-認識率
#1
#3
#10
#15
#16
#18
#19
#20
#41
#42
average
チャンネル数

【図１０】


100
98
96
94
92
90
88
認識率
#1
#3
#10
#15
#16
#18
#19
#20
#41
#42
話者
認識率
平均認識率
話者-認識率、平均認識率

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.